**Fuzzy Logic Applied Spatial Queries in Geographic Information Systems**

Angélica Félix de Castro and Trevor Martin

Departamento de Ciências Exatas e Naturais, Universidade Federal Rural do Semi-Árido, Mossoró, Brazil

Engineering Mathematics, University of Bristol, Bristol, UK

**Abstract:** In the recent years, the usage of Geographic Information Systems (GIS) has been rapidly increasing and it became the main tool for analyzing spatial data in unprecedented number of fields of activities. The processing time is important and currently uses more advanced applied mathematical and computer science knowledge: One of these mathematical theories is fuzzy logic. The fuzzy logic theory gives the possibility of enhancing spatial data management with the modelling of uncertainty. In this paper, the applicability of fuzzy set theory has been presented to respond spatial queries and the comparison performance between normal and spatial queries.

**Keywords:** Spatial Databases, Fuzzy Sets, Spatial Fuzzy, Performance

**Introduction**

Geographic Information System or Geographical Information System (shortly GIS) represents a powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world.

A GIS captures, stores, analyzes, manages and presents data that are linked to geographically referenced information. Geospatial data are special kind of data that must have some reference to locations on Earth and should depict some geographic phenomenon (known as themes), like land use, population density, weather etc., related to those locations. The geospatial data (Abel et al., 1999) have some special characteristics: (i) GEO-referencing, (ii) Spatial and non-spatial attributes, (iii) Spatial relationship and (sometimes) (iv) Temporal property (Mukherjee and Ghosh, 2011b).

In traditional GIS, there are quantitative operations. Further, the real world is more complex and full of uncertainties. Humans, however, often prefer a qualitative operation over an exact quantitative one. The properties of geospatial object and relationship among them may be ambiguous, which can be achieved by extending the standard map, overlaying operations to fuzzy maps and using fuzzy logic rather than crisp logic.

Uncertainty in geospatial data is often considered in the context of geographical information systems which enable a variety of operations and manipulation of spatial data. Fuzzy set and rough set theory has been used to represent geospatial data with uncertainty. It is common to use fuzzy sets in terrain modelling and triangulated irregular networks techniques.

This paper proposes the use of a fuzzy spatial technique to respond not exact spatial queries. The idea is to apply fuzzy logic in difficult queries to answer with normal SQL commands. The paper in organized as follows. Section 2 presents a theoretical framework background about GIS, Spatial Data, Fuzzy Sets and Fuzzy Spatial Data. The technologies and datasets used are explained in section 3. Section 4 describes the experimental results and the Section 5 presents the conclusion and future work.

**Theoretical Framework**

Geographic Information System or Geographical Information System (GIS) is a computer system for capturing, storing, checking and displaying data related to positions on Earth’s surface. GIS can show many different kinds of data on one map. This enables people to more easily see, analyze and understand patterns and relationships. With GIS technology, people can compare the locations of different things in order to discover how they relate to each other. For example, using GIS, the same map could include sites that produce pollution, such as gas stations and sites that are sensitive to pollution, such as wetlands. Such a map would help people determine which wetlands are most at risk.

GIS turns data, through analysis, into useful information. GIS stores spatial data (with its logically linked attribute information) in a GIS storage database,
where analytical functions are controlled interactively to generate the needed information products. Spatial data is raw data distinguished by the presence of a geographic link. In other words, the data is connected to a known place on the Earth, a true geographic reference. The features are viewed in the map - roads, likes, buildings - are ones commonly found in a GIS database as individual thematic layers. Linked to these geographic features and usually stored in a table format, is non-spatial information about them, data such as the name of the road, seasonal temperatures of the lake, owner of the building. These various characteristics applied to place are called attributes in GIS parlance and, in fact, it is the range and depth of these attributes that make spatial data such a powerful tool (Tomlinson, 2007).

Fuzzy set was introduced by Zadeh (1963). Nowadays, fuzzy sets theory and fuzzy logics are widely used for capturing the inherent vagueness that exists in real life applications. In classical set theory an element either belong to a set or does not belong to a set, in fuzzy set theory containment of an elements a set is represented with a certain degree. Formally, let \( U \) be the collection of elements called universe of discourse. A crisp subset \( A \) of \( U \) is collection elements \( U \) that is defined by a characteristic function \( \mu_A(u) \) that assigns any element \( u \in U \) to a value 1 if the element belongs to \( A \) and 0 if it does not belong to set \( A \). On other hand, for a fuzzy subset \( A \) of \( U \) is defined by a membership function \( \mu_A(u) \) for \( u \in U \). This membership function assigns any \( u \in U \) a value between 0 and 1 that represents degree to which and element of \( U \) belong to subset \( A \) of \( U \). For example, a classical set of real numbers greater than 6 can be expressed as \( A = \{ u | u > 6 \} \). For similar condition fuzzy set can be defined as \( A = \{ u \in U | \mu_A(u) > 0 \} \).

Currently, most of the GIS systems represent spatial datasets by crisp spatial objects. In reality, most spatial features have fuzzy characteristics. The crisp spatial data models fail to describe the spatial features correctly and may cause loss of information. In order to derive better resolution of spatial query, there is a need to describe and represent the fuzzy properties in spatial data sets and their relationships. Mukherjee and Ghosh (2011a) developed a Fuzzy Spatial Data Model, where apply fuzzy in spatial data.

**Datasets and Technologies used**

**Traffic Accidents in United Kingdom**

An accident is defined as a traffic accident if it occurs on a road or in a place to which the public have access. This can include footpaths and bridleways. Reported Road Casualties Great Britain (RCGB), formerly Road Casualties Great Britain (RCGB) and before that Road Accidents Great Britain (RAGB), is the official statistical publication of the UK Department for Transport (DfT) on traffic casualties, fatalities and related road safety data. This publication, first produced in 1951, is the primary source for data on road casualties in Great Britain. It is based primarily on police STATS19 data.

Data has been collected since 1926, in which year there were 4,886 fatalities in some 124,000 crashes. Between 1951 and 2006 a total of 309,144 people were killed and 17.6 million were injured in accidents on British roads. The highest number of deaths in any one year was 9,169 people in 1941 during World War II (DT, 2016).

In 1999, when Great Britain had the safest roads in Europe apart from Sweden, the government set a new national casualty reduction target, to be met by the year 2010. The target for 2010, compared to the average for the years 1994 to 1998, was a reduction of 40% in the number of people Killed or Seriously Injured (KSI) casualties, a reduction of 50% the number of children KSI casualties and a reduction of 10% in the rate of people slightly injured per 100 million vehicle kilometers. By 2009, the results were: killed or seriously injured 44% lower; children killed or seriously injured 61% lower and the slight casualty rate was 37% lower.

According with DT (2016), since 2010 until 2014 the official numbers of all accidents are: 2010 (154,414 accidents), 2011 (151,474), 2012 (145,571), 2013 (138,660) and 2014 (146,322).

**Hospitals in United Kingdom**

Healthcare in the United Kingdom is a devolved matter, meaning England, Northern Ireland, Scotland and Wales each have their own systems of publicly funded healthcare (UKHC, 2016).

The most recent comparison from the World Health Organization is now significantly out of date: in 2000, it ranked the provision of healthcare in the United Kingdom as fifteenth best in Europe and eighteenth in the world. In their 2014 edition, the Commonwealth Fund’s Mirror, Mirror on the Wall report, which ranks the top eleven first world healthcare systems, placed the United Kingdom as first overall taking first place in the following categories: Quality of Care (i.e., effective, safe, coordinated and patient-oriented subcategories), Access to care, Efficiency and Equity. The UK system had placed 2nd just four years previous in the 2010 report (Davis et al., 2014).

Today, there are 1022 hospitals in all United Kingdom.

**Used Datasets**

In this study, two datasets were used: (1) *Accidents2005_2014.csv*, where there is information about all STATS19 data (accident, casualties and vehicle...
tables) for 2005 to 2014 and (2) Hospital.csv with information about each hospital. These datasets were downloaded from the government site with official data of the UK (UKG, 2016).

Both datasets have attributes describing each entity. The table Accidents2005_2014 has 32 columns and 1,640,486 rows and the table Hospital has 17 columns and 1,022 rows.

Both datasets have Latitude and Longitude attributes and can be considered spatial data.

Applying specific UML to spatial data, we have the

**Fig. 1.**

### Used Technologies

After downloading the two official datasets with UK’s data (Accidents2005_2014 and Hospital), we converted both datasets into spatial database to do queries and view in spatial data. The Accidents2005_2014 has many important information as accidents severity, number of vehicles, number of casualties, local road, local authority and lighting conditions, among others.

To realize queries and view these data, we used three technologies to work with spatial data: PostgreSQL + PostGIS and Quantum GIS.

PostgreSQL is an Object-Relational Database Management System (ORDBMS) based on POSTGRES, Version 4.2, developed at the University of California at Berkeley Computer Science Department. PostgreSQL can be extended by the user in many ways, for example by adding new data types, functions, operators, aggregate functions, index methods and procedural languages (PostgreSQL, 2016).

PostGIS is an extension to the PostgreSQL system which allows GIS objects to be stored in the database (PostGIS, 2016). PostGIS is an open source, freely available and fairly Open Geospatial Consortium (OGC) compliant spatial database extender for the PostgreSQL. In a nutshell it adds spatial functions such as distance, area, union, intersection and specialty geometry data types to the database.

Quantum GIS (or QGIS) is a cross-platform free and open-source desktop GIS application that provides data viewing, editing and analysis capabilities. Similar to other software GIS systems QGIS allows users to create maps with many layers using different map projections (Quantum GIS, 2016).

Thus, with three technologies together, the queries and analysis were done and analyzed in Accidents2005_2014 and Hospital datasets.

### Queries and Analysis in Spatial Database: Integration of Imprecise Information using Fuzzy

This work was realized in a 2.2 GHz, Intel Core i5 machine with 8 GB RAM.

Firstly, we did the download of Accidents2005_2014 and Hospital datasets from official site of the UK. After, we understood the data each table and converted the file.csv (from site) to .shp (spatial format) using Postgis commands. It was possible because both tables have Latitude and Longitude columns.

After this conversion, we visualized the spatial data in QGis. Both spatial layers can be viewed in Fig. 2.

The queries realized are described below:

**Query 1: How Many Fatal Accidents Did Happen in UK from 2005 to 2014?**

Fatal accidents in Table Accidents lements severity, number of vehicles, number of casualties, local road, local authority and lighting conditions, among others.

To realize queries and view these data, we used three technologies to work with spatial data: PostgreSQL + PostGIS and Quantum GIS.

PostgreSQL is an Object-Relational Database Management System (ORDBMS) based on POSTGRES, Version 4.2, developed at the University of California at Berkeley Computer Science Department. PostgreSQL can be extended by the user in many ways, for example by adding new data types, functions, operators, aggregate functions, index methods and procedural languages (PostgreSQL, 2016).

PostGIS is an extension to the PostgreSQL system which allows GIS objects to be stored in the database (PostGIS, 2016). PostGIS is an open source, freely available and fairly Open Geospatial Consortium (OGC) compliant spatial database extender for the PostgreSQL. In a nutshell it adds spatial functions such as distance, area, union, intersection and specialty geometry data types to the database.

Quantum GIS (or QGIS) is a cross-platform free and open-source desktop GIS application that provides data viewing, editing and analysis capabilities. Similar to other software GIS systems QGIS allows users to create maps with many layers using different map projections (Quantum GIS, 2016).

Thus, with three technologies together, the queries and analysis were done and analyzed in Accidents2005_2014 and Hospital datasets.

**Queries and Analysis in Spatial Database: Integration of Imprecise Information using Fuzzy**

This work was realized in a 2.2 GHz, Intel Core i5 machine with 8 GB RAM.

Firstly, we did the download of Accidents2005_2014 and Hospital datasets from official site of the UK. After, we understood the data each table and converted the file.csv (from site) to .shp (spatial format) using Postgis commands. It was possible because both tables have Latitude and Longitude columns.

After this conversion, we visualized the spatial data in QGis. Both spatial layers can be viewed in Fig. 2.

The queries realized are described below:

**Query 1: How Many Fatal Accidents Did Happen in UK from 2005 to 2014?**

Fatal accidents in Table Accidents2005_2014 are integer numbers. The column Accident_S is divided in: Fatal, Serious and Slight. The Fatal category is the number 1. To respond this question, we used normal commands SQL.

SQL Query: Select* from "Accidents2005_2014" Where "Accident_S" = 1
Answer: 21382 fatal accidents in nine years
Processing Time in PostgreSQL: 4297 ms

**Query 2: How Many Accidents Did Occur in Bristol city from 2005 to 2014?**

In the same way, the cities are represented in the table as numbers. The Bristol code is 601 in Accidents2005_2014 table. Again, we use normal commands SQL to respond this query.

SQL Query: Select * from "Accidents2005_2014" Where "Local_Auth" = 601
Answer: 11859 accidents in Bristol during nine years
Processing Time in PostgreSQL: 2470 ms

**Query 3: How Many Serious Accidents Did Occur in London on Sundays?**

In the table, serious accident (Accident_S column) is represented by number 2; the London city (Local_Auth column) is number 570 and day of week is 1 to represent Sundays.

SQL Query: Select * from "Accidents2005_2014" Where "Accident_S" = 2 AND "Local_Auth" = 570 AND "Day_of_Wee" = 1
Answer: 26 accidents
Processing Time: 557 ms
Fig. 1: Spatial UML applied in Accidents2009_2014 and Hospital datasets.

Fig. 2: Layers with data of: (a) Accidents2005_2014; (b) Hospital and (c) Layers Hospital and Accidents2005_2014 together.
Query 4: The Accident Occurred on 10.07.2008, at 06:00 am, with Serious Victims, what the Ten Nearest Hospitals to Take the Victims?

This query is a spatial query because there is important spatial factor (nearest hospital). The Postgis was used to respond this question because it has specific spatial commands. The clause \( \text{ST Distance (geom1, geom2)} \) was utilized to calculate the distance between the local of accident and the nearest hospital. LIMIT 10 limits the query in 10 hospitals.

```
SELECT g1."Latitude", g1."Longitude", g2."Organisa_4" as NameHospital FROM "Accidents2005_2014" AS g1, "Hospital" AS g2 WHERE "Accident_S" = 2 AND "Date" like '10/07/2008' AND "Time" like '06:00' ORDER BY ST_Distance(g1.the_geom,g2.the_geom) LIMIT 10
Answer: 10 rows with 10 nearest hospital
```

The Fig. 3 shows the local where the accident happened and ten nearest hospitals. Quantum GIS was used to visualize it: UK map and after zoom, it was possible to see the local of accident (red star) and the nearest hospitals (black points).

Sometimes, the spatial query is not trivial and it is not easy to respond. As an example, consider the following query, Q5, which involves uncertainty and there are not exact values.

Query 5: Find Accidents with High Number of Casualties and What Hospitals Nearest Each Accident

This question is not possible to be solved with SQL commands because we do not have numeric values or exact terms to elaborate the query. In addition, the query Q5 has some fuzzy linguistic terms, like “high” and “nearest”. For example, hospital nearest accidents may include hospitals that are neighbors of the accident or those which are neighbors the accident with some membership degree. Then, fuzzy spatial data model was adapted and used to solve this query.

Firstly, classes and attribute sets of a class may be fuzzy, i.e., existence of a class have possibility to the model and existence of attribute in class have a possibility. In this work, the class Accidents2005_2014 is a fuzzy class but Hospital is a normal class.

Another important level of fuzziness is related to the attributes with fuzzy values. For this a keyword FUZZY is placed in front of the attribute. In the class Accidents2005_2014, the Number_of_Casualties attribute may have fuzzy values in terms of linguistic expression like low, medium, high and/or very high.

The association between the classes represents the relationship between the objects. For example (refer Fig. 4) an association NearestOf exists between Hospital and Accidents2005_2014 classes. The association exist with possibility degree (deg Nearest).

Figure 4 represents a fuzzy geospatial data model that defines the structure of the datasets Accidents2005_2014 and Hospital.

The Number_of_Casualties attribute of class Accidents2005_2014 is a specific case because an exact definition does not exist. What is high number of victims? Five victims? Ten victims? And a low number? One victim? Two victims? Thus a fuzzy set theory has been applied to capture the fuzzy values. The Number_of_Casualties (NC) membership function \( \mu_{\text{NC}}(n) \) is defined by trapezoid and triangular membership function (Fig. 5). Both membership functions are extensively used in real time implementations due to their simple formula. For same reason these membership functions have been used here.
Describing the function $\mu_{NC}(n)$ of accidents with respect to Number of Casualties (NC), the universe of discourse of NC is 0 to 100. The membership functions are described using trapezoidal and triangular membership functions (Fig. 6 and Fig. 7).

The Number of Casualties = low and Number of Casualties = very high are given with trapezoidal membership function and Number of Casualties = medium and Number of Casualties = high are given with triangular membership function.

Suppose data provider A1 has data of Accidentes2005_2014 and H1 has data of Hospital. Firstly, a SQL query found the all accidents with number of casualties.

```
SELECT "Accident_I", "Local_Auth", num as NameHospital, "City"
FROM "Accidents2005_2014" AS g1, "Hospital" AS g2
WHERE "Accident_I" like <Accident_Index>
ORDER BY ST_Distance(g1.the_geom,g2.the_geom) LIMIT 3
```

The result based on fuzzy data model is given in Fig. 8. Analyzing Fig. 08, some conclusions were obtained:

- The accidents A2 and A4 were close each other, where the hospitals nearest A2 were H1, H2 and H3
- About accident A4, could lead the victims to H2, H3, H7, H8, H9
- The accident A3 presented three hospital close it (H4, H5, H6)
- The accident A5 was the most difficult: There were 68 victims and no hospital close. The nearest hospital were H4, H5, H6
Table 1: Number of casualties value calculation

| Accident | Accident_index | Local_authority_(Highway) | Number of casualties | μ
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>2006460127192</td>
<td>Kent</td>
<td>20</td>
<td>0.00</td>
</tr>
<tr>
<td>A2</td>
<td>2006610080206</td>
<td>Monmouthshire</td>
<td>41</td>
<td>0.30</td>
</tr>
<tr>
<td>A3</td>
<td>201142I121804</td>
<td>Essex</td>
<td>51</td>
<td>0.80</td>
</tr>
<tr>
<td>A4</td>
<td>2014530209023</td>
<td>Gloucestershire</td>
<td>54</td>
<td>0.95</td>
</tr>
<tr>
<td>A5</td>
<td>200743N002017</td>
<td>Buckinghamshire</td>
<td>68</td>
<td>0.35</td>
</tr>
<tr>
<td>A6</td>
<td>2013460234852</td>
<td>Kent</td>
<td>70</td>
<td>0.25</td>
</tr>
<tr>
<td>A7</td>
<td>201106X047581</td>
<td>Salford</td>
<td>87</td>
<td>0.00</td>
</tr>
<tr>
<td>A8</td>
<td>201441000489</td>
<td>Hertfordshire</td>
<td>93</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 2: Accidents with high number of casualties and the nearest hospital

<table>
<thead>
<tr>
<th>Accident</th>
<th>Accident_index</th>
<th>Local_Authority_ (Highway)</th>
<th>Number of casualties</th>
<th>Hospital nearest</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>2006610080206</td>
<td>Monmouthshire</td>
<td>732</td>
<td>H1) Partnerships in Care Llanarth, H2) Dilke Memorial Hospital, H3) Lydney and District Hospital in Lydney</td>
</tr>
<tr>
<td>A3</td>
<td>201142I121804</td>
<td>Essex</td>
<td>463</td>
<td>H4) Saffron Walden Community Hospital in Saffron Walden, H5) Herts and Essex Community Hospital in Bishop's Stortford, H6) Rosie Hospital in Cambridge</td>
</tr>
<tr>
<td>A4</td>
<td>2014530209023</td>
<td>Gloucestershire</td>
<td>624</td>
<td>H7) Vale Hospital in Dursley, H8) Vale Community Hospital, H9) Westridge</td>
</tr>
<tr>
<td>A5</td>
<td>200743N002017</td>
<td>Buckinghamshire</td>
<td>463</td>
<td>H4) Saffron Walden Community Hospital in Saffron Walden, H5) Herts and Essex Community Hospital in Bishop's Stortford, H6) Rosie Hospital in Cambridge</td>
</tr>
</tbody>
</table>

Fig. 6: Trapezoidal membership functions

Fig. 7: Triangular membership functions
In this way, the query not-exact “Find accidents with high number of casualties and what hospitals nearest each accident” was been answered using fuzzy logic. Thus, fuzzy spatial data model is able to represent the information which is closed to the given condition. A realistic and flexible result is obtained, which may be important for real world decision support system.

**Experimental Results**

Five queries were executed in spatial data Accidents2005_2014 and Hospital. The queries Q1, Q2 and Q3 were normal queries, without spatial component; Q4 was a spatial query and Q5 was a query which the answer was based in fuzzy logic.
Table 3: Queries with number of rows and time processing (ms)

<table>
<thead>
<tr>
<th>Query</th>
<th>Rows</th>
<th>Time processing (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 (normal)</td>
<td>21382</td>
<td>4297</td>
</tr>
<tr>
<td>Q2 (normal)</td>
<td>11859</td>
<td>2470</td>
</tr>
<tr>
<td>Q3 (normal)</td>
<td>26</td>
<td>557</td>
</tr>
<tr>
<td>Q4 (spatial)</td>
<td>10</td>
<td>29172</td>
</tr>
</tbody>
</table>

The amount the rows and processing time were registered in each query. The Table 3 and Fig. 9 show these values. Q5 was not registered because it was a kind different query and processing, not only using SQL commands.

Analyzing the Fig. 9, some considerations are important:

- Q1, Q2 and Q3 showed a pattern behaviour: Higher number of rows, greater processing time
- The spatial query Q4 had the lowest number of rows and the largest processing time
- Even with the lowest number of rows as final result, a spatial query takes more time than a normal query
- Q3 processed 26 rows in 557 ms and Q4 processed 10 (spatial) rows in 29172 ms. Processing time Q4 was 52 times greater than the processing time spent by Q3

Thus, spatial queries need more time to do the processing.

Conclusion

This paper has presented concepts of Geographical Information System, Spatial Database, Spatial Queries, Fuzzy Sets and Fuzzy Spatial Data Model. We downloaded two official datasets from UK government, one with all accidents occurred in country between 2005 and 2014 and another with information about hospital in England. Both datasets were useful to view, understand and analyse the accidents during this period and the appropriate hospital.

To realize data analysis, we used normal SQL commands and spatial SQL commands in PostGIS. The result of queries could be visualized using PostgreSQL (tables) or QGIS (points, lines and polygons). In non-exact queries, we used fuzzy sets and fuzzy spatial data model to respond and the final answer could be visualized in QGIS. Based on fuzzy set and fuzzy set theory operation, the information uncertainty was managed efficiently. The integration and query processing based on fuzzy spatial data models give more realistic information that can be used in various decision making processes.

We examined our performance results and the conclusion is that spatial queries consume more time than normal queries. The explanation is that in a spatial query the processing happens in non-spatial attributes (tables) and spatial data (points, lines and polygons). It needs processing both information.

About the accidents during 2005 and 2014, the most serious accidents (with and without casualties) occurred in South and Southeast of UK. Safety and prevention measures should be adopted in these locations.

As future work the idea is use fuzzy in scalable in spatial databases and explore the advantages to use fuzzy in spatial databases.

Acknowledgement

This study received financial support from the Science without Borders Program and the Brazilian National Council for Scientific and Technological Development (CNPq). The author is grateful to the Universidade Federal Rural do Semi-Arido (Brazil) and the University of Bristol (UK).

Funding Information

This study received financial support from the Science without Borders Program and the Brazilian National Council for Scientific and Technological Development (CNPq).
Author’s Contributions

Angélica Félix de Castro: She did the study and she wrote the paper.

Trevor Martin: This research was carried out during a visit to the Intelligent Systems Lab, University of Bristol, hosted by professor Trevor Martin. He provided the entire structure and study laboratory.

Ethics

Authors will address any ethical issues that may arise after the publication of this manuscript.

References


Quantum GIS, 2016. QGIS - The leading open source desktop GIS. https://www.qgis.org/en/site/about/
